Traps: Identifying Mortality-related Cliques in Comorbidity Network

Pankush Kalgotra, PhD

Assistant Professor of Business Analytics

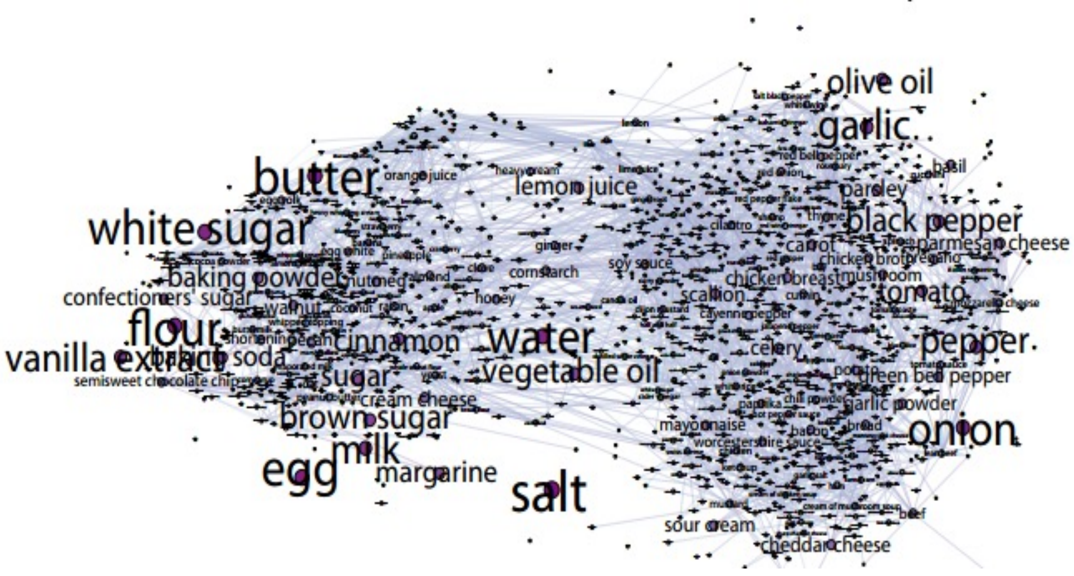
Department of Business Analytics and Information Systems

Harbert College of Business

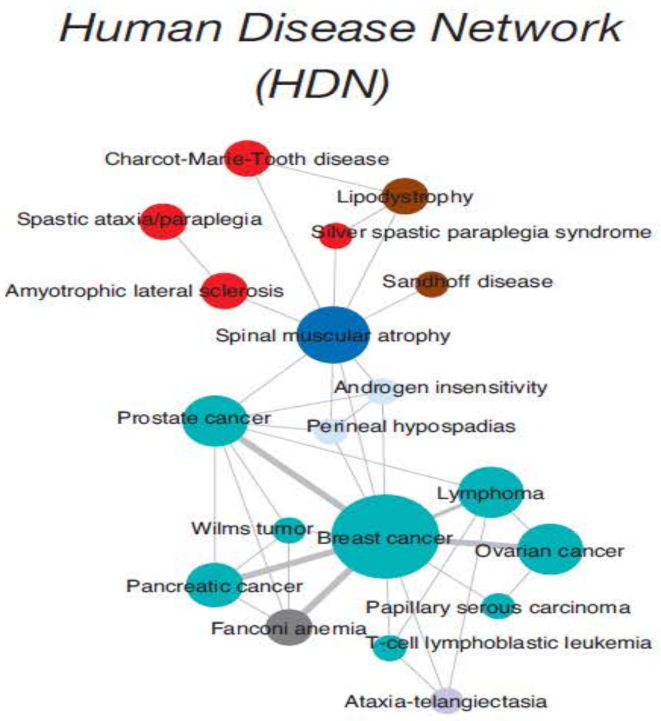
Auburn University

AUBURN
UNIVERSITY

Network Science



- Network method – nodes and relationships
  - Explicit Networks (Facebook Network)
  - Implicit/Inferred Networks (Comorbidity Network)



- Implicit Network
- Relationships are inferred – similarity index

Network analytics research

Link predictions

Information diffusion

Sample the network

Impact of network on its nodes, which are internal to the network

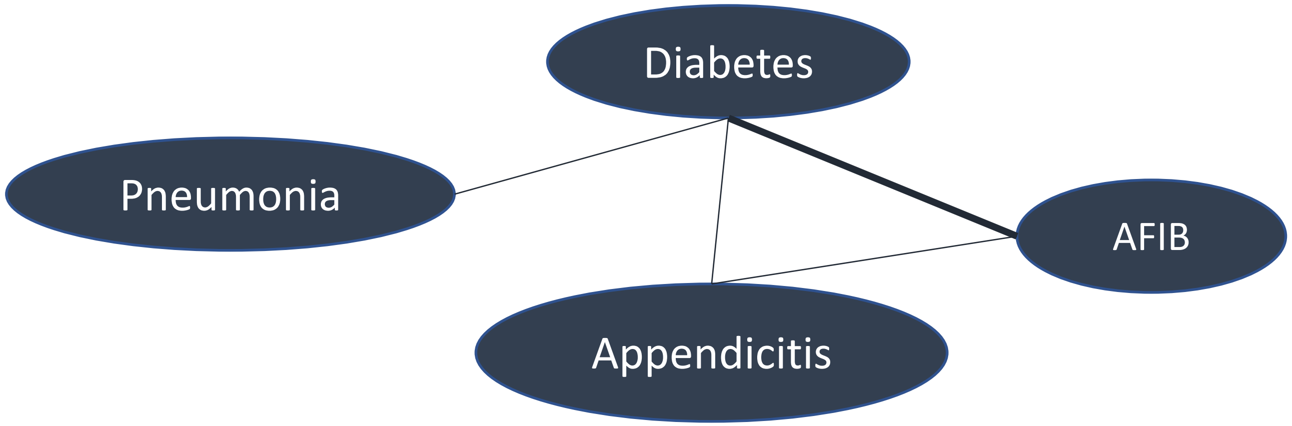
Recommendation engine

Limited work on understanding the impact of network on exogenous outcomes

Comorbidity Network

- Network of diseases – comorbidity network
  - diseases linked to each other based on co-occurrences in patients (comorbidity)

  - Node: a disease (diagnosis)
  - Edge: comorbidity/co-occurrence



Dataset

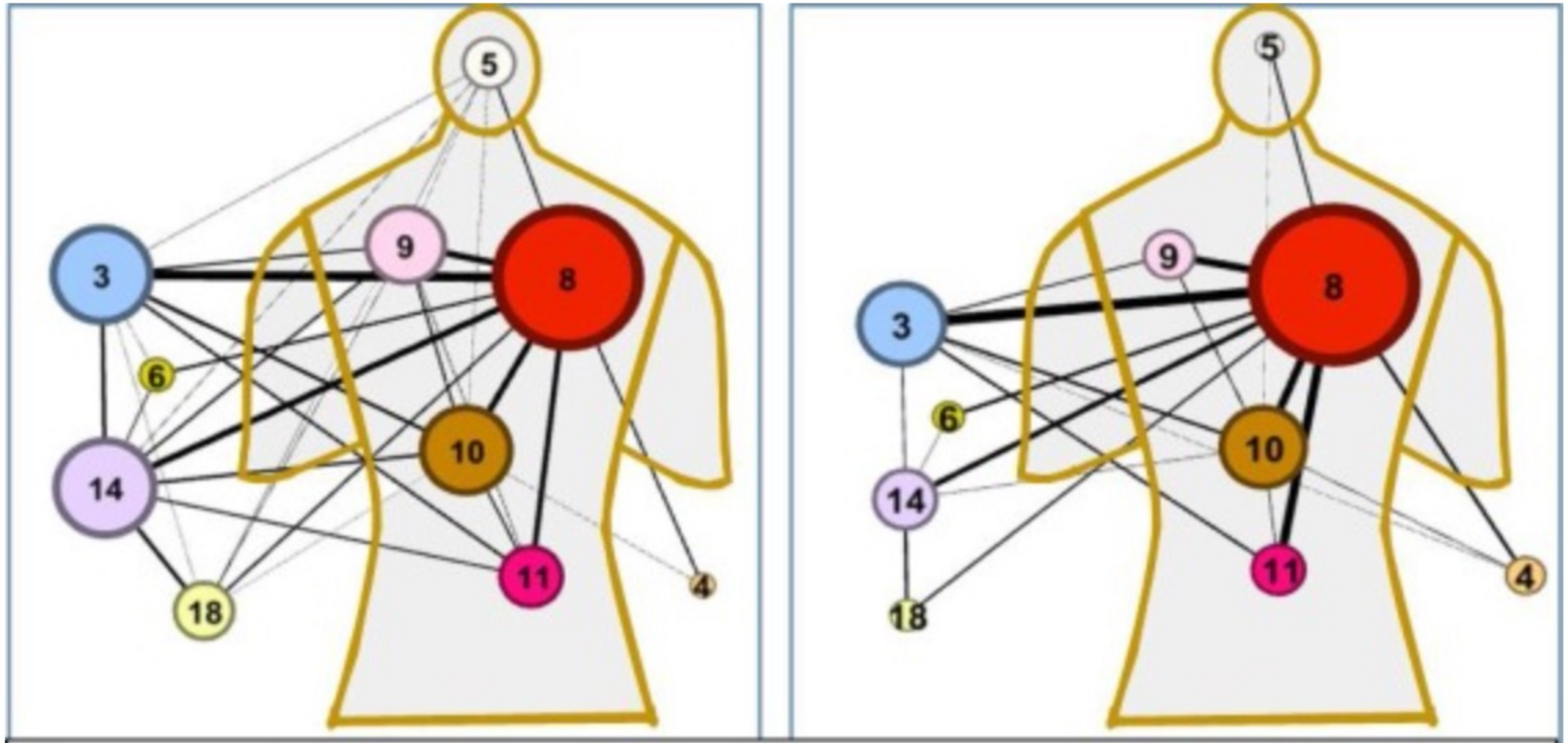
- Electronic medical records (EMR) – Patient history
- Cerner database
  - more than 50 million patients' visits

- Disease is measured in ICD-9-CM codes

428 Heart failure
- 428.0 Congestive heart failure, unspecified
- 428.1 Left heart failure
- 428.2 Systolic heart failure
  - 428.20 Systolic heart failure, unspecified
  - 428.21 Acute systolic heart failure
  - 428.22 Chronic systolic heart failure
  - 428.23 Acute on chronic systolic heart failure

Descriptive Analytics

- Kalgotra, P., Sharda, R., & Croff, J. M. (2017). Examining health disparities by gender: a multimorbidity network analysis of electronic medical record. *International journal of medical informatics*, *108*, 22-28.



2a. Female Organ Comorbidity Network 2b. Male Organ Comorbidity Network

Descriptive Analytics

Kalgotra, P., Sharda, R., & Croff, J. M. (2020). Examining multimorbidity differences across racial groups: a network analysis of electronic medical records. *Nature Scientific Reports*, 10(1), 1-9.

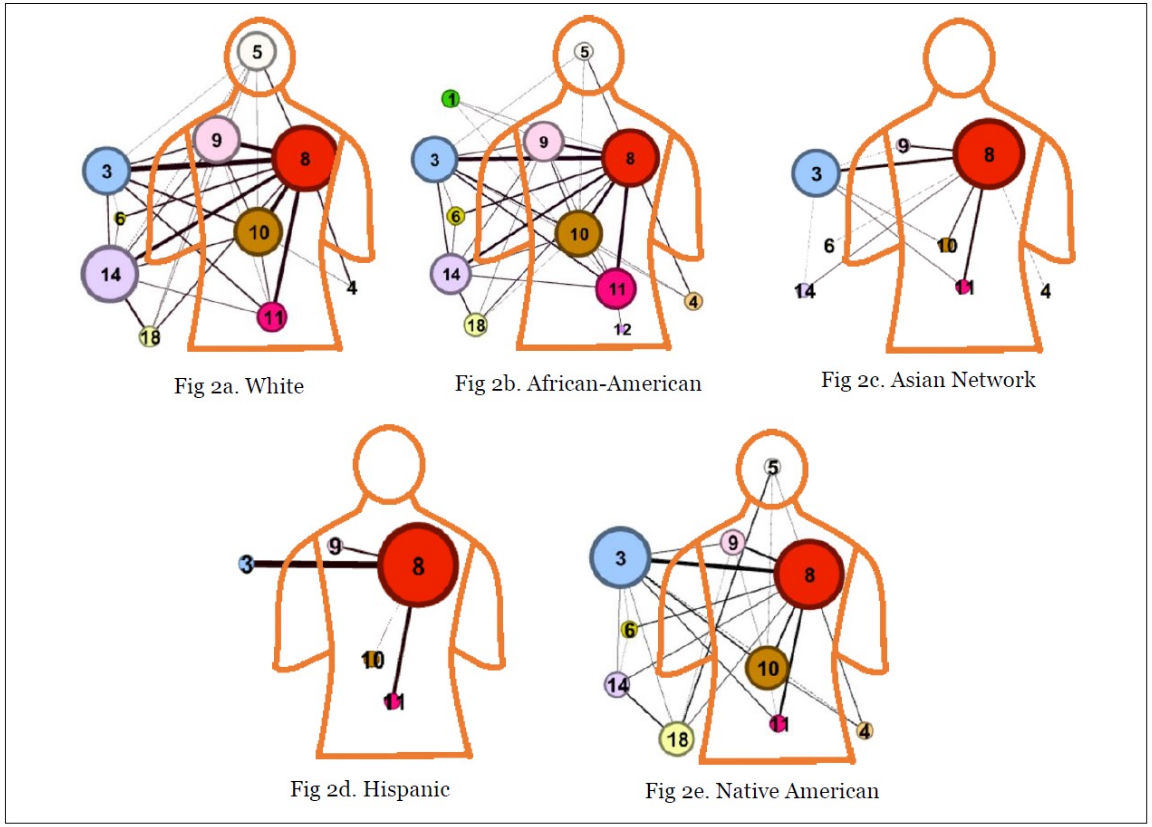


Fig 2. Multimorbidity Network by Race at Organ System Level

Predictive Analytics

Kalgotra, P., & Sharda, R. (2021). **When will I get out of the Hospital? Modeling length of stay using comorbidity networks.** *Journal of Management Information Systems*, 38(4), 1150-1184.

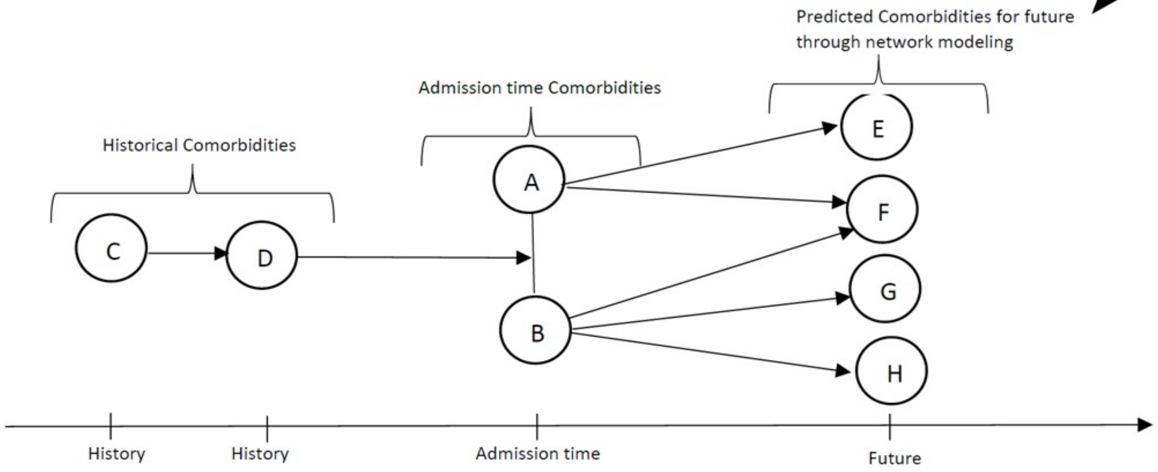
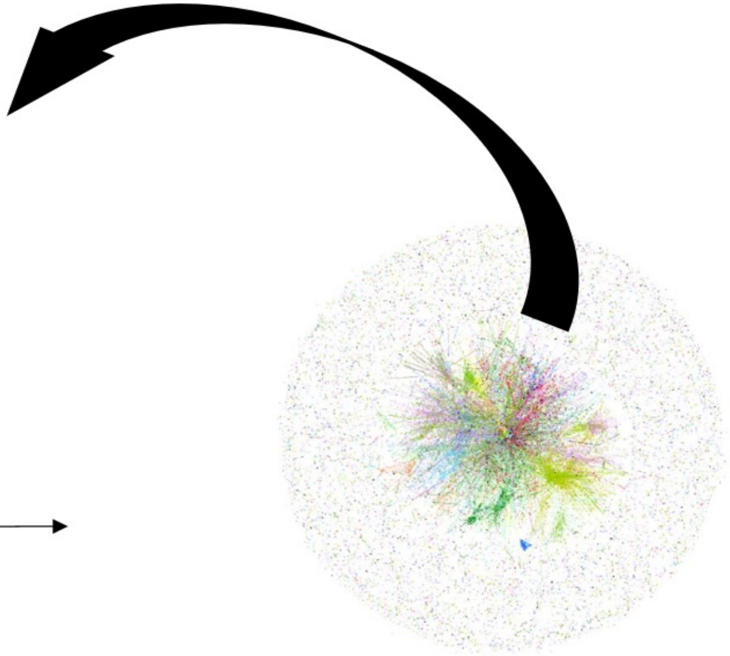


Figure 1. A patient's timeline and comorbidities



$$\begin{bmatrix} LOS & Pchar_1 & \dots & Pchar_k & d_{o1} & d_{o2} & \dots & d_{on} & \begin{matrix} d_{h1} \\ d_{p1} \end{matrix} & \begin{matrix} d_{h2} \\ d_{p2} \end{matrix} & \begin{matrix} d_{h3} \\ d_{p3} \end{matrix} & \begin{matrix} \dots \\ \dots \end{matrix} & \begin{matrix} d_{hn-1} \\ d_{pn-1} \end{matrix} & \begin{matrix} d_{hn} \\ d_{pn} \end{matrix} \end{bmatrix}$$

Predictive Analytics

Kalgotra, P., Sharda, R., Parasa, S. (2023). **Quantifying Disease-Interactions through Co-occurrence Matrices to Predict Early Onset Colorectal Cancer.** *Decision Support Systems*. In Press.

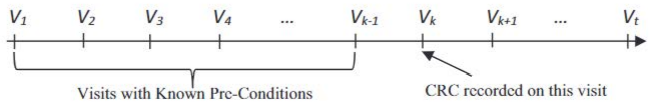


Fig. 1. A patient timeline.

$$C = \begin{bmatrix} 0 & c_{12} & \cdots & c_{1h} \\ c_{21} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_{h1} & \cdots & & 0 \end{bmatrix}$$

$$SumC = \sum_e \sum_f c_{ef}$$

$$AverageC = \frac{\sum_e \sum_f c_{ef}}{\frac{r(r-1)}{2}}$$

- Created classification models for early-onset colorectal cancer (EoCRC)
- Proposed novel disease interaction variables
- Created score variables to estimate the risk of EoCRC
- Models can help in prescribing colonoscopy screening

Modeling results.

Model	Accuracy (%), SD	AUC, SD	Sensitivity, (%), SD	Specificity, (%), SD
Model 1: Age	67.5, 0.5	0.73, 0.01	71.1, 1.8	64.0, 1.8
Model 2: Age, *Dem*	68.5, 0.6	0.74, 0.01	72.1, 1.7	64.9, 1.7
Model 3: Age, *Dem*, *Sym*	72.5, 0.6	0.80, 0.01	73.6, 1.1	71.5, 1.5
Model 4: Age, *Dem*, *DI*	70.5, 0.6	0.77, 0.01	73.4, 0.8	67.6, 1.1
Model 5: *Sym*	67.5, 0.6	0.73, 0.01	60.6, 2.2	**74.4, 2.6**
Model 6: *DI*	67.0, 0.5	0.72, 0.01	65.7, 1.0	68.3, 1.4
Model 7: Age, *Dem*, *Sym*, *DI*	**73.2, 0.4**	**0.81, 0.01**	**75.3, 0.6**	71.1, 0.8

Traps: Identifying Mortality-related Cliques in Comorbidity Network

Pankush Kalgotra and Ramesh Sharda

Introduction

- Hidden combination of actions/activities impact the outcomes

- Combination of
  - actors in a movie impact movie success
  - authors in a paper impact citation count
  - drugs impact a patient’s outcomes
  - diseases impact a patient’s outcomes

Objectives

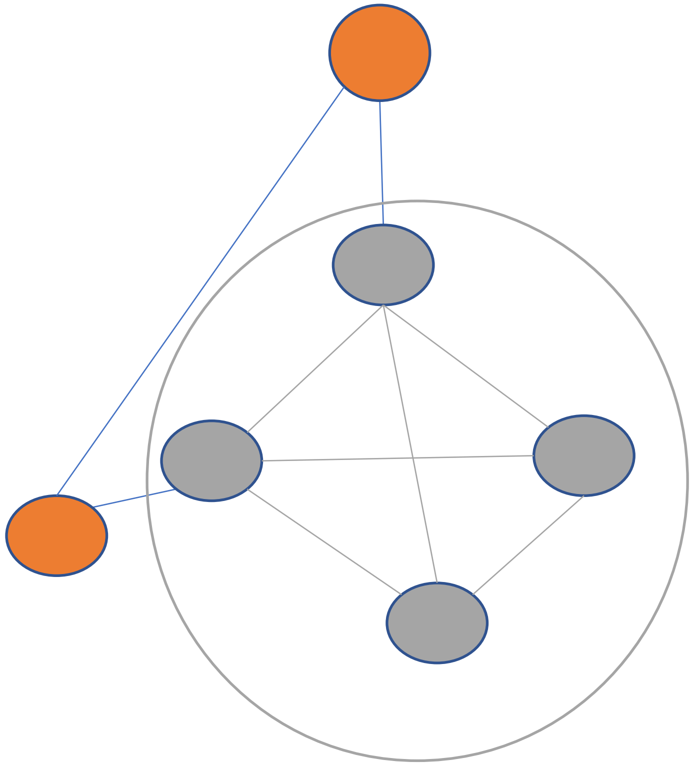
- To model combinations
- To develop a method to find hidden combinations related to an outcome in a situation

Solution

- Network Theory to model combinations
  - Latent Network
  - Identifying Cliques in the network, which are highly related to an outcome

Clique

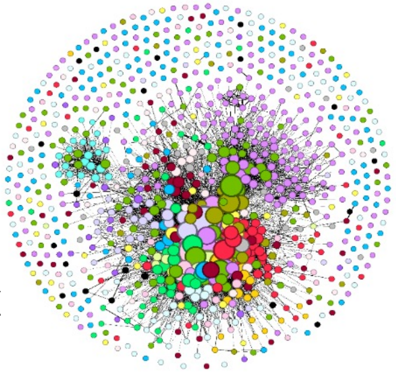
- Sub-network: all individuals know each other (complete structure)
- Implications on the performance -trust, norms and obligations (Coleman, 1988)
- Provan and Sebastian (1998) - cliques are positively related to the network effectiveness



Demonstration: Mortality related Cliques

- Causes of mortality – diseases, actually multiple diseases
- Direct and indirect interactions of diseases may be related to mortality
- Which combinations of diseases are critical?

Demonstration: Mortality related Cliques



- Network of diseases – comorbidity network
- Diseases linked to each other based on co-occurrences in patients– Salton Cosine Index

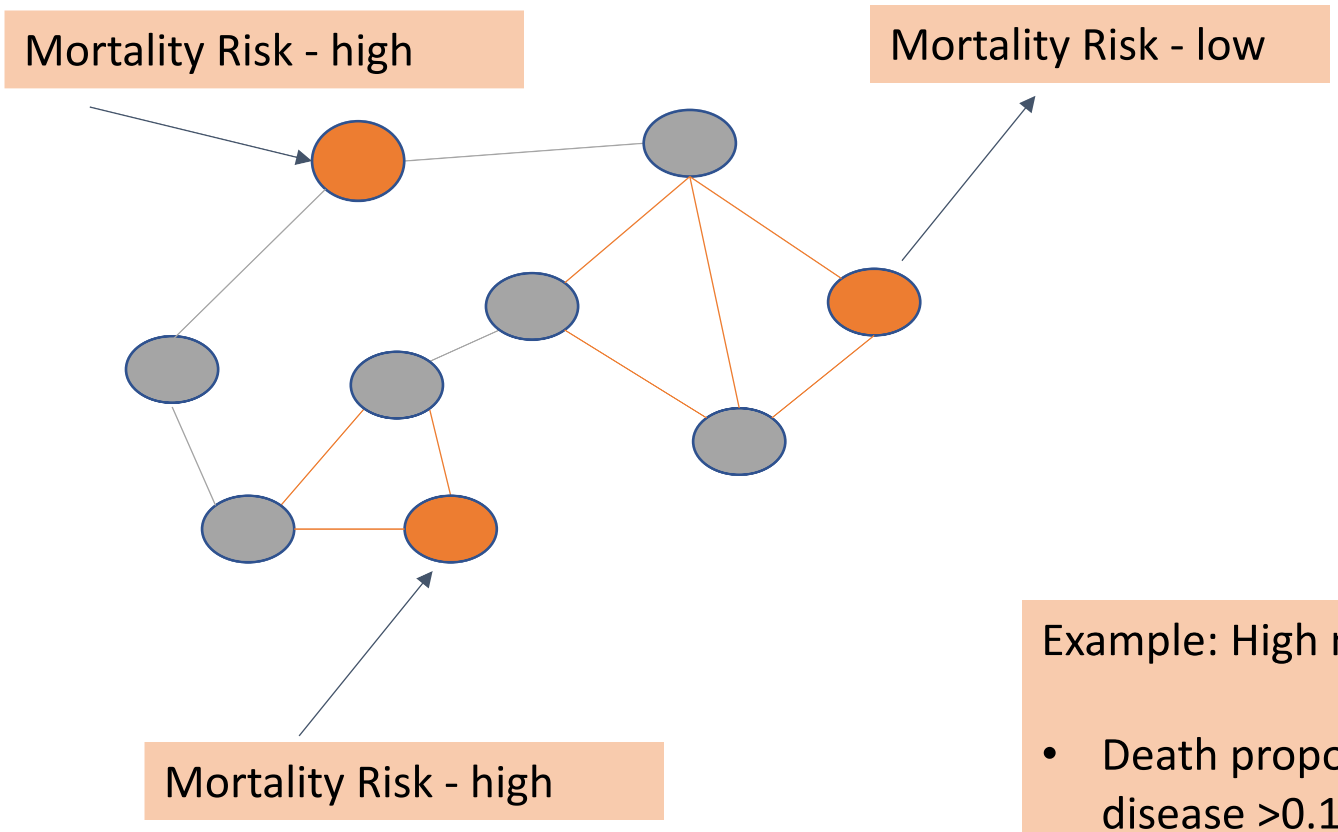
$$SCI_{ij} = \frac{(c_{ij})}{\sqrt{(c_i * c_j)}}$$

- Dataset: Cerner Electronic Medical Records (2000-2015)
  - 24.7 M Patients
  - 1,043 unique diagnoses & symptoms
  - 23,313 edges
  - 4.46% dense

$$D_{nXn} = \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & \dots & & 0 \end{bmatrix}$$

Algorithm

- Step 1: Identify high risk diseases

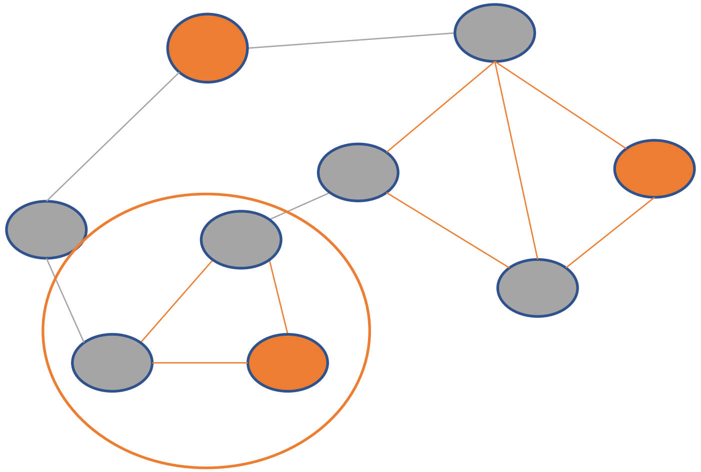


Example: High mortality risk if:

- Death proportion due to the disease >0.1

Algorithm

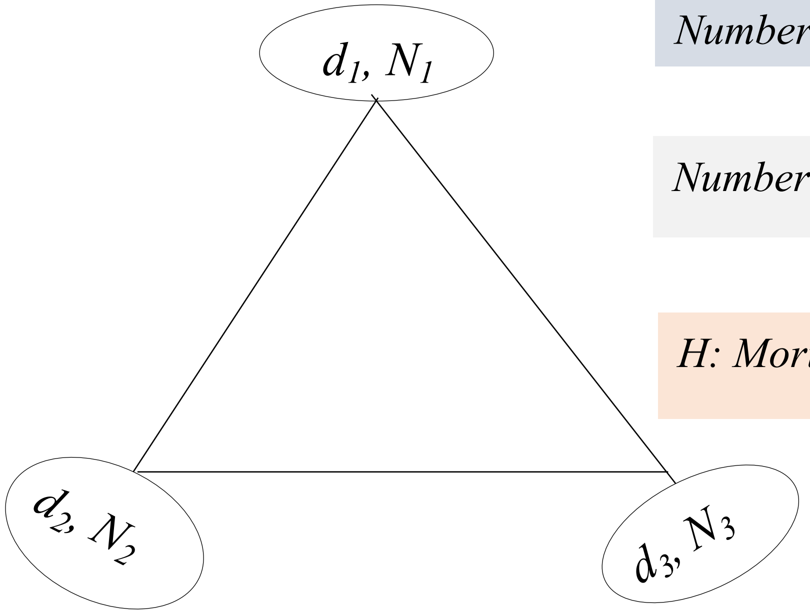
- Step 2: Find the maximal clique using Clustering Coefficient



$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

- Step 3: Compute and compare the mortality risk with and without a clique
  - *H: MortalityRisk (N123) ≥ MortalityRisk (N1- N123)*

Clique in a comorbidity network

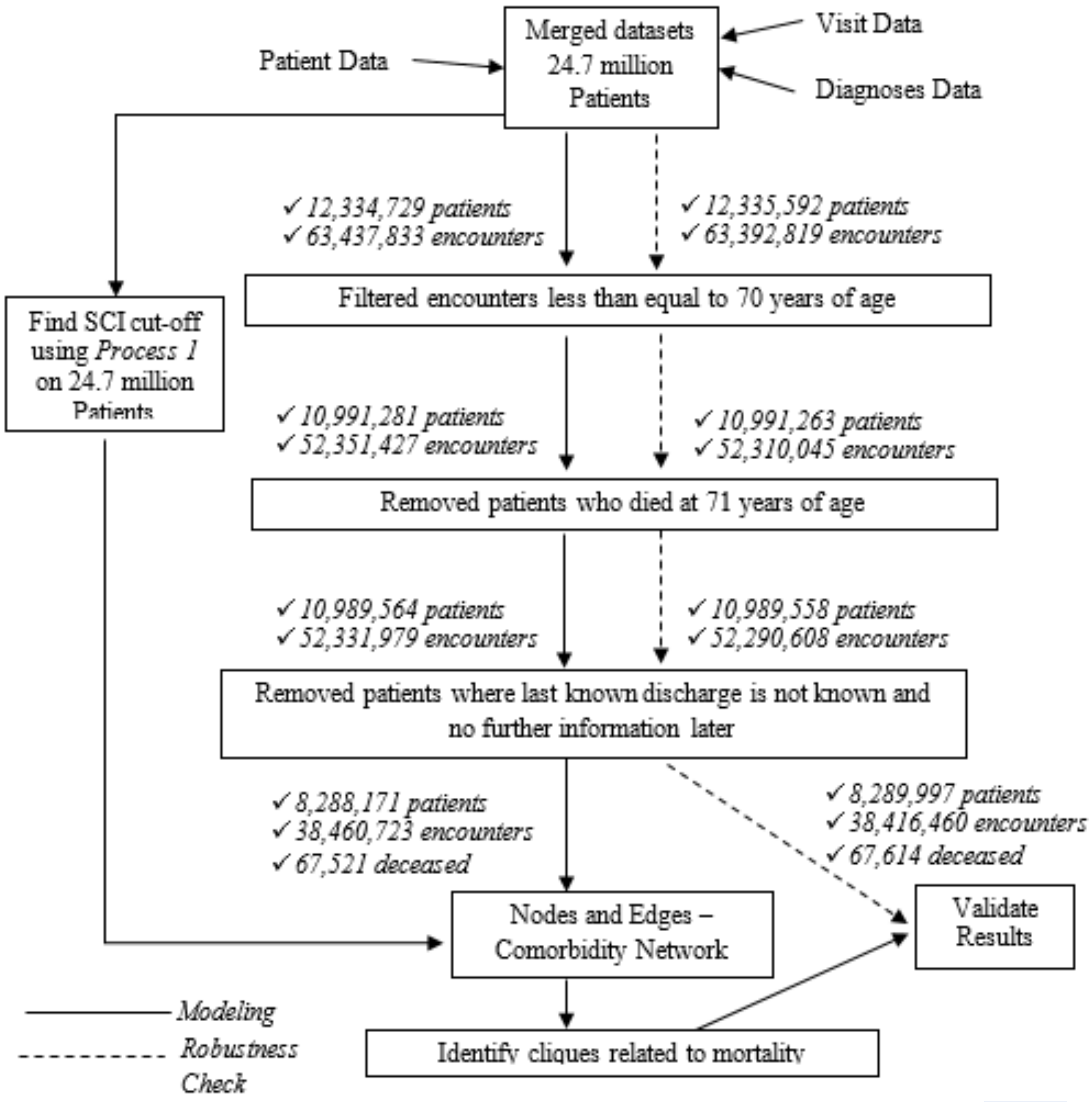


Base Disease – d_1

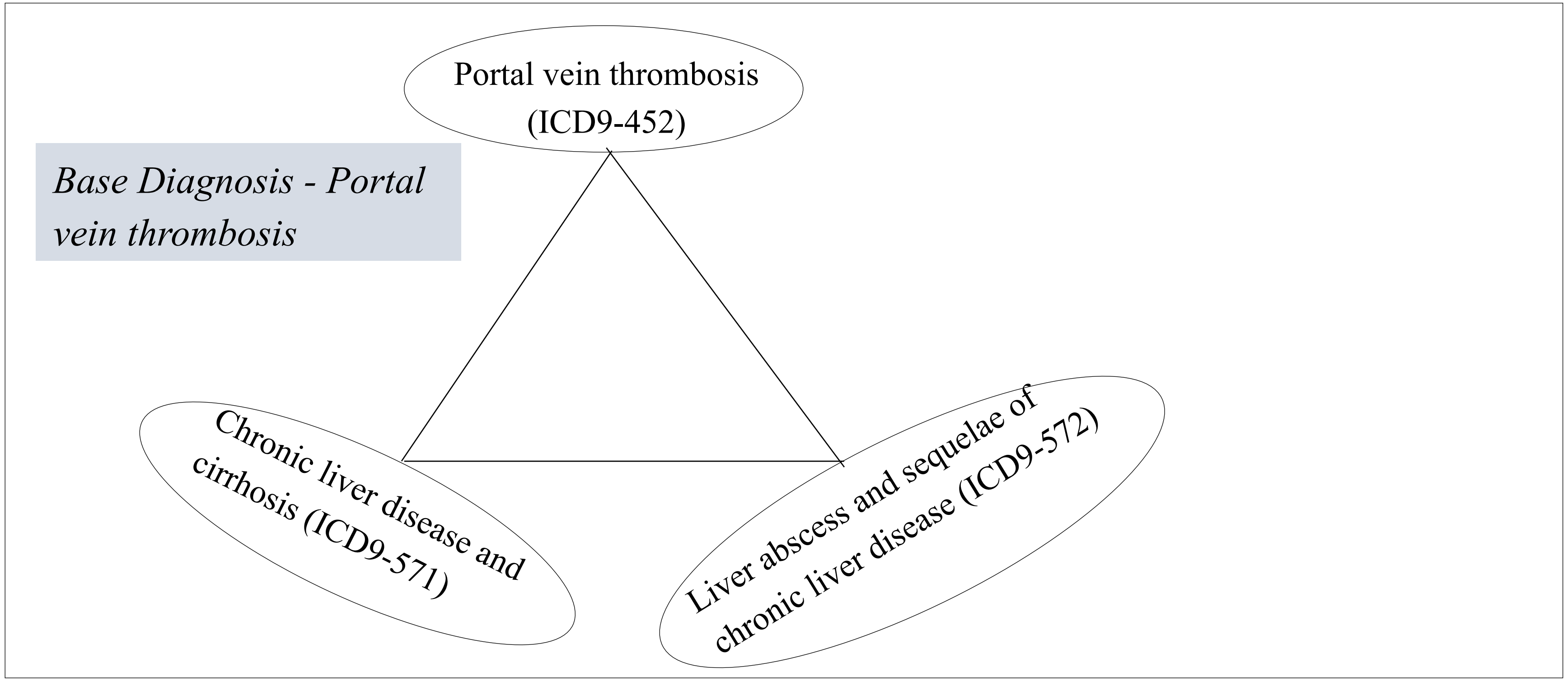
Number of patients diagnosed with $d_1 = N_1$

Number of patients with all three diagnoses $= N_{123}$

H: MortalityRisk $(N_{123}) \geq$ MortalityRisk $(N_1 - N_{123})$


Merged datasets
24.7 million
Patients
Visit Data
Patient Data
Diagnoses Data
✓ 12,334,729 patients
✓ 63,437,833 encounters
✓ 12,335,592 patients
✓ 63,392,819 encounters
Find SCI cut-off using Process 1 on 24.7 million Patients
Filtered encounters less than equal to 70 years of age
✓ 10,991,281 patients
✓ 52,351,427 encounters
✓ 10,991,263 patients
✓ 52,310,045 encounters
Removed patients who died at 71 years of age
✓ 10,989,564 patients
✓ 52,331,979 encounters
✓ 10,989,558 patients
✓ 52,290,608 encounters
Removed patients where last known discharge is not known and no further information later
✓ 8,288,171 patients
✓ 38,460,723 encounters
✓ 67,521 deceased
✓ 8,289,997 patients
✓ 38,416,460 encounters
✓ 67,614 deceased
Nodes and Edges – Comorbidity Network
Validate Results
Identify cliques related to mortality
Modeling
Robustness Check

Results



A clique/triangle of three diseases with their joint impact on mortality

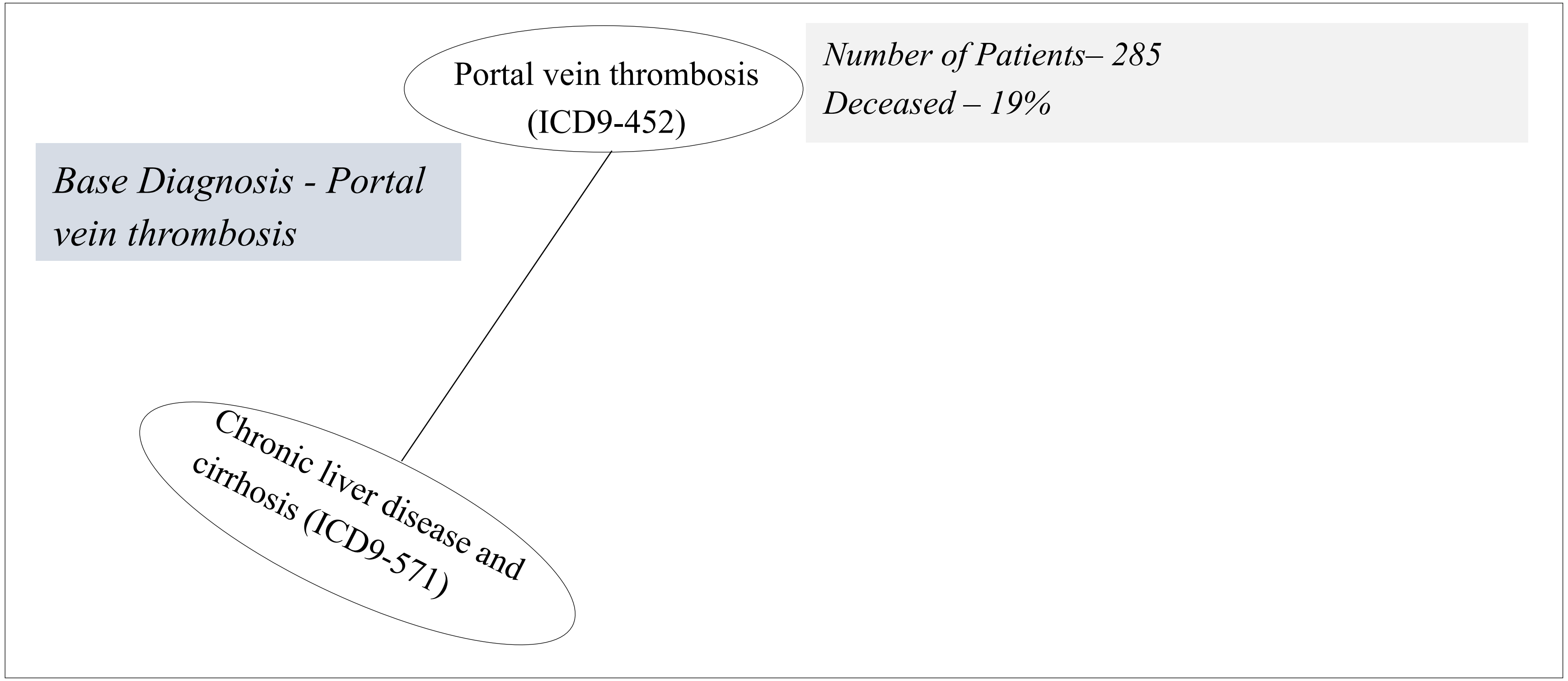
Results

Portal vein thrombosis (ICD9-452)

Number of Patients– 836
Deceased – 16%

Base Diagnosis - Portal vein thrombosis

Results

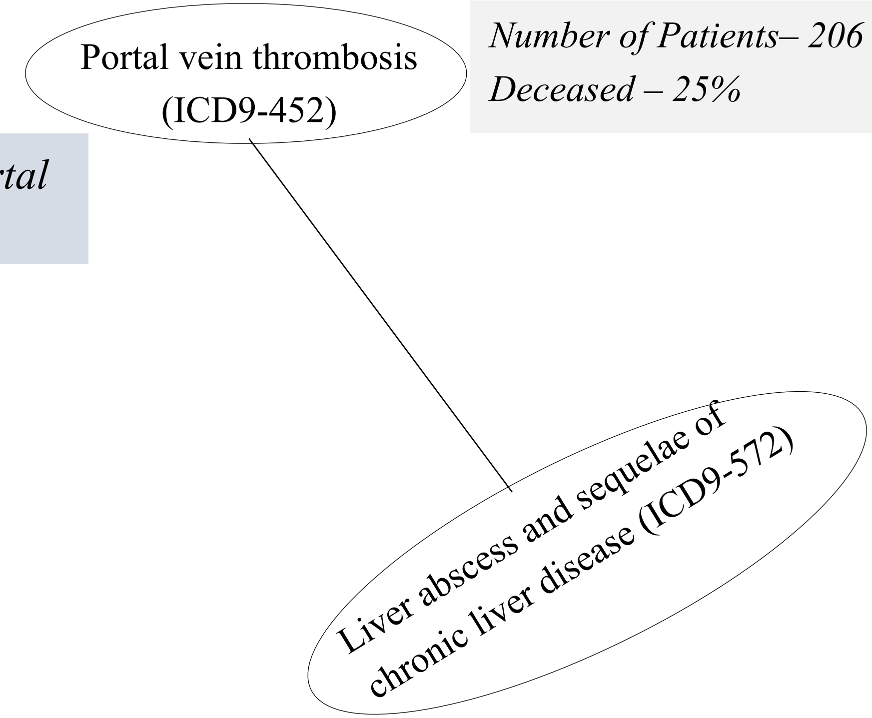

Portal vein thrombosis (ICD9-452)
Number of Patients– 285
Deceased – 19%
Base Diagnosis - Portal vein thrombosis
Chronic liver disease and cirrhosis (ICD9-571)

Results

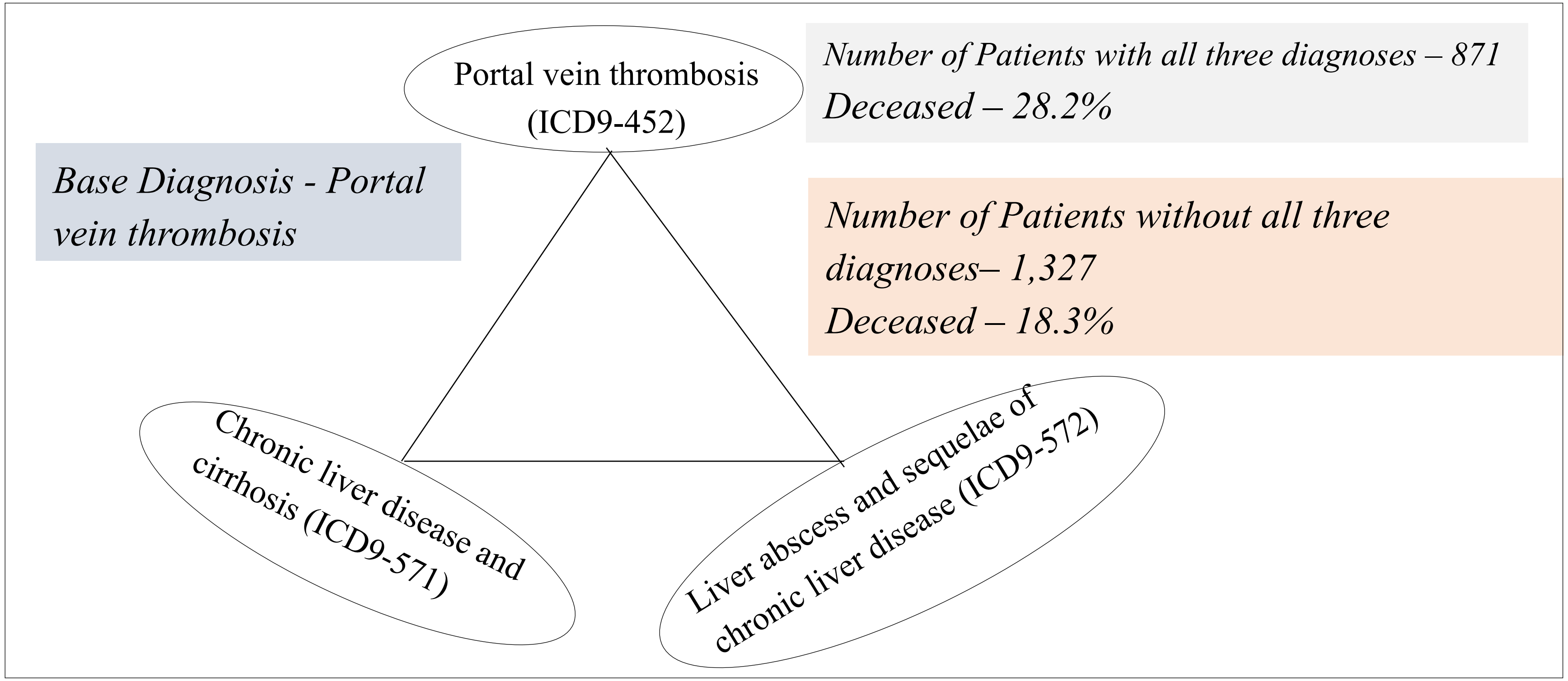
Number of Patients– 206

Deceased – 25%

Base Diagnosis - Portal vein thrombosis



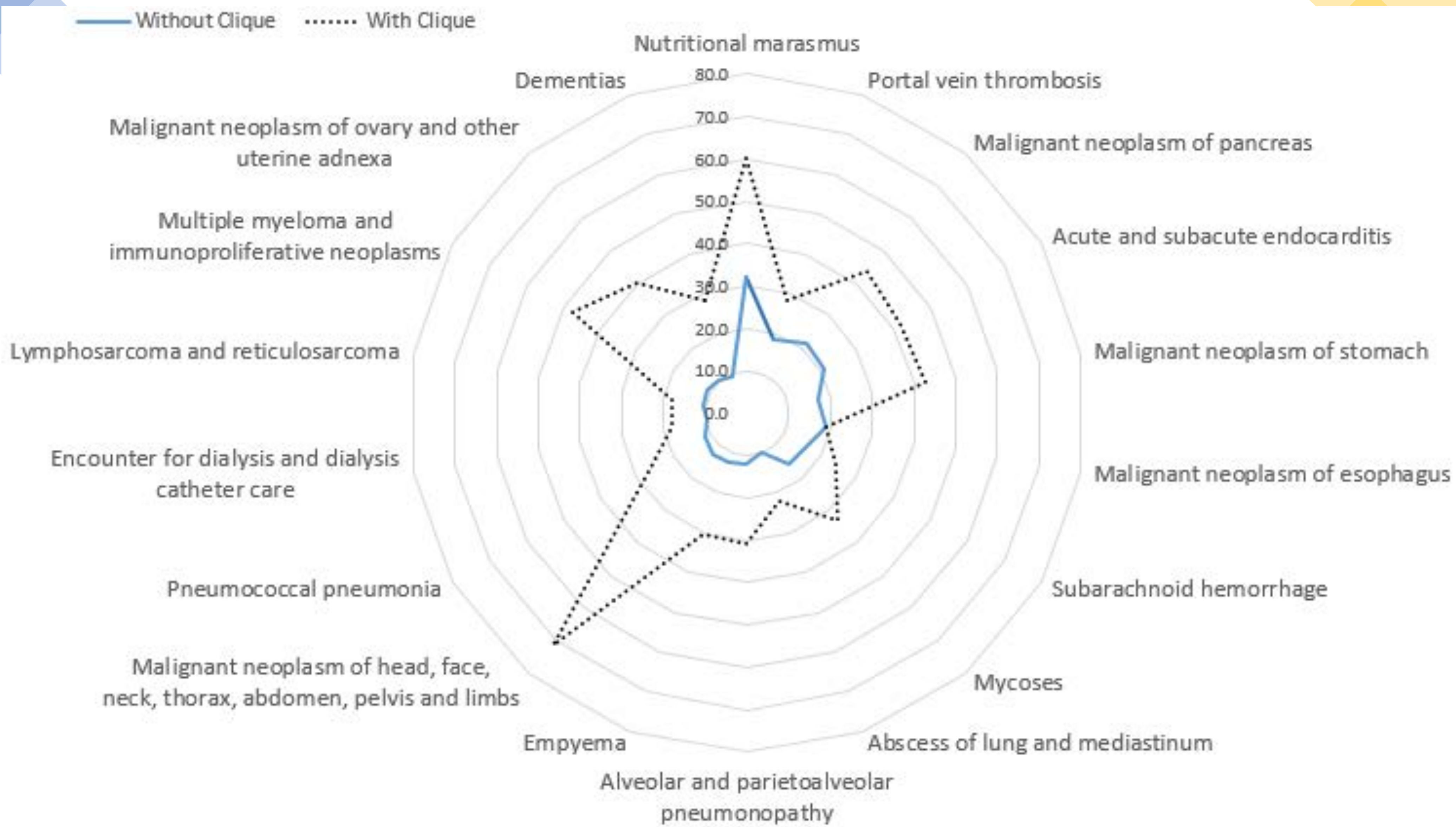
Results



A clique/triangle of three diseases with their joint impact on mortality

Results

Mortality rate with and without cliques						
ICD-9	Description	Clique Size	Patient count w/o Clique	Mortality rate w/o a clique (%)	Patient count with Clique	Mortality rate with clique (%)
261*	Nutritional marasmus	13	3,725	32.0	30	60
452*	Portal vein thrombosis	3	1,327	18.3	871	28.2
157*	Malignant neoplasm of pancreas	7	5,808	21.5	73	43.8
421*	Acute and subacute endocarditis	15	3,760	21.2	67	41.8
151*	Malignant neoplasm of stomach	3	2,968	17.1	427	42.9
150	Malignant neoplasm of esophagus	3	2,772	18.8	905	19.0
430*	Subarachnoid hemorrhage	4	5,520	16.0	192	24.0
117*	Mycoses	16	6,558	15.6	21	33.3
513*	Abscess of lung and mediastinum	3	1,513	9.8	659	22.2
516*	Alveolar and parietoalveolar pneumonopathy	6	4,013	12.0	201	30.8
510*	Empyema	9	3,985	12.1	151	30.5
195*	Malignant neoplasm of head, face, neck, thorax, abdomen, pelvis and limbs	12	6,096	12.6	14	71.4
481*	Pneumococcal pneumonia	8	4,098	11.6	160	26.3
V56*	Encounter for dialysis and dialysis catheter care	9	2,074	9.5	628	18.2
200*	Lymphosarcoma and reticulosarcoma	5	3,845	10.5	375	18.1
203*	Multiple myeloma and immunoproliferative neoplasms	12	5,843	10.8	19	47.4
183*	Malignant neoplasm of ovary and other uterine adnexa	7	6,929	10.2	142	40.1
290*	Dementias	8	4,305	9.5	152	28.3


Without Clique
With Clique
Nutritional marasmus
Portal vein thrombosis
Malignant neoplasm of pancreas
Acute and subacute endocarditis
Malignant neoplasm of stomach
Malignant neoplasm of esophagus
Subarachnoid hemorrhage
Mycoses
Abscess of lung and mediastinum
Alveolar and parietoalveolar pneumonopathy
Empyema
Malignant neoplasm of head, face, neck, thorax, abdomen, pelvis and limbs
Pneumococcal pneumonia
Encounter for dialysis and dialysis catheter care
Lymphosarcoma and reticulosarcoma
Multiple myeloma and immunoproliferative neoplasms
Malignant neoplasm of ovary and other uterine adnexa
Dementias
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0

Discussion/Contributions

- Structural property to theorize important health outcome
- Preemptive action required if a clique is found
- Identified maximal clique but some diagnoses might be redundant

- Clique/combination of event as a trap
- Outcome related cliques - algorithmic contribution
- Theorize phenomenon using network properties
- Method applicable to problems where latent interactions affect an outcome

Current status

- Presented at Informs conference
- Creating an optimization model

Acknowledgements

This work was conducted with data from the Cerner Corporation's Health Facts database of electronic medical records provided by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Cerner Corporation.

Email: pzk0031@auburn.edu

Thank You